

**IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK**

STATE OF NEW YORK, et al.,

Plaintiffs,

NEW YORK IMMIGRATION
COALITION, et al.,

Consolidated Plaintiffs

v.

UNITED STATES DEPARTMENT OF
COMMERCE, et al.,

Defendants.

Civil Action No. 1:18-cv-2921-JMF

CORRECTED AFFIDAVIT OF DR. WILLIAM P. O'HARE

I. QUALIFICATIONS

1. I was retained in this litigation to analyze the impact that the expected reduction in self-response rates during the 2020 Census will have on the accuracy of the 2020 Census. My analysis was offered in response to Dr. Abowd's assertion that reduced self-response rates resulting from the addition of a citizenship question will not impact the undercount. I demonstrate that this assertion is incorrect by showing that, historically, decreases in self-response rates strongly correlate with increased undercount rates, and that a citizenship question, by reducing self-response, is likely to increase undercount in 2020. A copy of the rebuttal report that I submitted in this case is PX-320, and the errata to that report is PX-334.

2. I have more than forty years of experience using Census data in a variety of professional settings, including experience in non-profits, philanthropy, state government, and university settings. Since 1987, I have worked at the Population Reference Bureau, the

University of Louisville, and the Annie E. Casey Foundation.

3. I have published many articles in scientific journals and written many books and book chapters based on Census Bureau data. I have authored more than a dozen monographs on subjects such as the well-being of children, poverty, and minorities in America. I have also made many presentations using Census data at professional conferences. While serving as the Director of the KIDS COUNT project within the Annie E. Casey Foundation from 1993 to 2006, I supervised extensive use of Census Bureau data related to measuring and reporting on the well-being of children.

4. I have been a member of the Population Association of America, the Southern Demographic Association, and the American Statistical Association since the 1970s. I served on the Board of Directors and was President of the Southern Demographic Association. I was a founding member and I served on the Board of Directors for the International Society of Child Indicators.

5. I served on the Census Bureau advisory committees focused on the 2000 and 2010 Censuses. From 1995 to 2001, I was a representative from the American Statistical Association on the Census Bureau's Professional Advisory Committee. From 2008 to 2011, I was a representative from the Association of Public Data Users on the Census Bureau's 2010 Census Advisory Committee. In addition, I was awarded a National Science Foundation/American Statistical Association/Census Bureau Research Fellowship where I worked at the Census Bureau from 2011 to 2013.

6. I have a Bachelor of Science Degree from Michigan State University in Multi-Disciplinary Social Science. I have a master's degree from Michigan State University in Multi-Disciplinary Social Science, I have a Ph.D. from Michigan State University in Sociology.

7. A copy of my C.V. is PX-335. Based on my experiences, I believe I am well-qualified to offer an expert opinion on the consequences of the anticipated decrease in self-response on the 2020 Census on undercount rates. I offer these opinions to a reasonable degree of scientific certainty.

II. INTRODUCTION

8. The Census Bureau expects the addition of the citizenship question on the 2020 Census to lead to a reduction of at least 5.1 percentage points in self-response rates for households with one or more noncitizens. The 5.1 percentage point decline in self-response rates for households with one or more noncitizens refers to a decrease in the share of those households completing a questionnaire in the first phase of the Census data collection operation.

9. My analysis demonstrates the close connection between lower self-response rates and higher net undercount rates and thus the inaccuracy of the position that Dr. Abowd articulated in his report of September 21, 2018 that a differential self-response decline, i.e. a reduction of self-response concentrated in one demographic groups such as non-citizens, will not result in a differential undercount, i.e. an increase in undercounting of that same demographic group. It is my opinion, based on the empirical statistical relationships from the 1990, 2000, and 2010 Decennial Censuses, as discussed in my analysis, that the Census Bureau's expected decrease of at least 5.1 percentage points in the self-response rates for households with at least one noncitizen because of the citizenship question will increase the net undercount and omission rates for people living in those households. There is no evidence in the 1990, 2000 or 2010 Censuses that the Census Bureau can mitigate low response rates to avoid a net undercount.

A. Key Concepts and Terms

10. Before presenting the analysis and results, it is important to define some key

concepts and terms used in my analysis.

11. In its simplest form, the U.S. Decennial Census can be thought of as having two major phases; 1) a self-response phase and 2) a Non-Response Follow Up (NRFU) phase. The self-response phase consists of households returning the questionnaire that was mailed to them from the Census Bureau (in 2020, for the first time in the decennial census, the self-response will include internet and telephone responses). Several weeks after census day (April 1) the second phase of the Census begins, and households that did not return a completed Census questionnaire are visited by a Census enumerator to gather the information needed for the Census. This is referred to as Non-Response Follow Up (NRFU) operations by the Census Bureau, and also includes the use of administrative records and third-party data to provide data for occupied housing units. While a few people belatedly self-respond during the NRFU phase, and there are other Census operations (like update leave and update enumerate) that are outside of these two operations, the bulk of Census respondents are captured in the self-response and NRFU operations of the Census.

12. Self-response rates reflect the percent of households that return the census questionnaire. In this analysis, self-response is measured by two closely related indicators: Mail Return Rates and Mail Response Rates, both of which are calculated by the Census Bureau. Detailed descriptions of the Mail Return Rates and the Mail Response Rates are provided by the U.S. Census Bureau. *See 2014 Planning Database (PX-345)*, at p. 61. In simple terms, the Mail Return Rate is the percentage of census questionnaires that were returned from occupied households. Mail Response Rates are the percentage of census questionnaires that were returned from all households whether they were occupied or not. The Mail Return Rate is used by Dr. Abowd to measure self-response. Where available, my analysis uses Mail Return Rate to

measure self-response rates. However, the Mail Return Rate was not available in the 1990 Census, so I use the Mail Response Rate to reflect the self-response rate in 1990.

13. While self-response modes other than mail, such as the internet and telephone, will be available during the 2020 Census, Mail Response and Mail Return Rates from prior censuses are the best indicator of self-response. Moreover, the Census Bureau has estimated that the self-response rate in the 2020 Census will be lower than the self-response rate in previous Censuses, suggesting that internet and telephone response options will not be effective at improving self-response compared to past Censuses.

14. Net undercount and omissions rates are both measures of Census accuracy, but they capture different parts of Census accuracy. According to the U.S. Census Bureau, “omissions are people who should have been enumerated in the United States Census but were not.” *See 2010 Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States*, 2012 (PX-267), at 12. Erroneous enumerations are people that are double counted, counted in the wrong place, and those inappropriately included in the Census – like foreign tourists. Imputed persons are those added to Census count based on evidence they exist, including persons added from a housing unit that looks occupied but where there is no self-response, and no one responds to an enumerator.

15. The net undercount is the balance between omissions and those included erroneously and those imputed. If the number of omissions is higher than the number of erroneous inclusions and whole person imputations, there is a net undercount. If the number of erroneous inclusions and whole person imputations is larger than the number of omissions, there is a net overcount.

16. Both omissions and erroneous enumerations are calculated by the Census Bureau

using the Dual-Systems Estimation method. While there is admittedly some uncertainty in these rates, they are the best estimates available, and appropriate for the analysis that I conducted.

17. Omissions can offer insights into populations missed by the Census that are masked by net undercount rates. For example, if 10% of Hispanics in a state are missed, while an equal number of Non-Hispanic Whites are double counted, the net undercount would be zero, but that does not reflect the fact that a large number of Hispanics were missed. In the 2010 Census there were 10,042,000 erroneous enumerations, 5,993,000 whole-person imputations and 15,999,000 omissions. *See* PX-267, at Table 3.

18. It is important to understand the net undercount often masks the extent to which specific groups were missed in the Census and it is worth noting that even when a net undercount for a group is zero, there are often omissions. For example, the net undercount of Asians in the 2010 Census was essentially zero, but there was an omissions rate of over 5 percent for Asians in the 2010 Census. Moreover the net undercount for children age 0-4 in the 2010 Census was 4.6 percent, but the omissions rate was 10.3 percent. *See* O'Hare, *Presentation on undercount of young children at Census Bureau/Children's Leadership Council meeting*, 2018 (PX-336).

19. Undercounts have sometimes been reported as a negative number by the Census Bureau and sometimes as a positive number by the Census Bureau. For this analysis, I refer to net undercounts as a positive number and net overcounts as a negative number. Measuring net undercounts here as a positive number makes the correlations and the figures easier to interpret.

B. Correlations

20. Much of my analysis relies on correlation coefficients to show relationships between two variables such as self-response rates and net undercount rates. More specifically, I use the Pearson Product-Moment Correlation Coefficient. This is probably the

most widely used correlation calculation.

21. There are three dimensions of correlation coefficients: direction, magnitude, and statistical significance. Direction is indicated by a positive or a negative sign. A positive correlation indicates a higher value on one variable is associated with a higher value in the other variable. The relationship between height and weight reflects a positive correlation, i.e. taller people usually weigh more. A negative sign indicates that a higher value on one variable is associated with a lower value in the other variable. The relationship between exercise and obesity reflects a negative correlation, i.e. people who exercise more are less likely to be obese.

22. The magnitude of a correlation coefficient varies from 0 to 1. A magnitude of zero means no relationship between the two variables and a value of 1 means a perfect correlation between the two variables. The higher the magnitude or value of the correlation coefficient the stronger the relationship between the two measures being examined.

23. Statistical significance testing is done to assess how likely the observed results are due to chance. Researchers use different levels of statistical significance depending on the analysis. The standard used by the Census Bureau in its publications and on its website is 0.10, which means if something is statistically significant the results would occur by chance alone less than one time out of ten. This is a commonly used benchmark in social science research. Another way of saying this is with a 0.10 level of significance we can be ninety percent confident the observed results reflect a real or true relationship between two measures.

24. All the correlation coefficients deemed statistically significant in my analysis are significant at the 0.10 level or higher (higher level of significance) – the same level that the Census Bureau uses. This means the observed results would happen by chance alone less than one time in ten. The ninety percent confidence level is a minimum. Most of the correlations in

my analysis are statistically significant at a much higher level. Since the observed results are unlikely to be due to chance, it is highly likely that the correlation coefficient reflects a real relationship, and these are not random results.

25. Since I am only interested in knowing if a correlation coefficient is negative and statistically significant, I use a one-tailed test of significance. A one-tailed test implies we are only interested in seeing if the correlation is negative and statistically different from zero. This contrasts to a two-tailed test which would tell us if the correlation was statistically different from zero in either direction that is positive or negative

26. This approach is a commonly used practice in social science research. Statistical significance is largely determined by the size of the correlation coefficient and the number of observations upon which the correlation is based. Higher magnitude and more observations lead to a higher-level statistical significance.

III. Self-Response and Census Accuracy

27. In this section, I examine the relationship between self-response rates and Census accuracy, as measured by net undercount and omissions rates. The 1990, 2000, and 2010 Census provide statistical data that can be used to examine this relationship from an empirical perspective. The Census net undercount and omissions rates are taken from Dual-Systems Estimates method (based on a Post-Enumeration Survey). The sample size for the Post-Enumeration Survey (PES) varied over the Census years. Specifically, the sample size of the 2010 PES was smaller than the sample size used in 2000 and, consequently, the estimates have higher standard errors. Nevertheless, the data published by the Census Bureau are the best available. Moreover, the data are sufficient to detect the relationship between levels of self-response and Census accuracy. The consistency of correlations across censuses, demographic

groups and states indicate the association between self-response rates and Census accuracy is a robust relationship. Relationships are shown graphically as well as statistically, as it may be easier to grasp a relationship from a visual presentation.

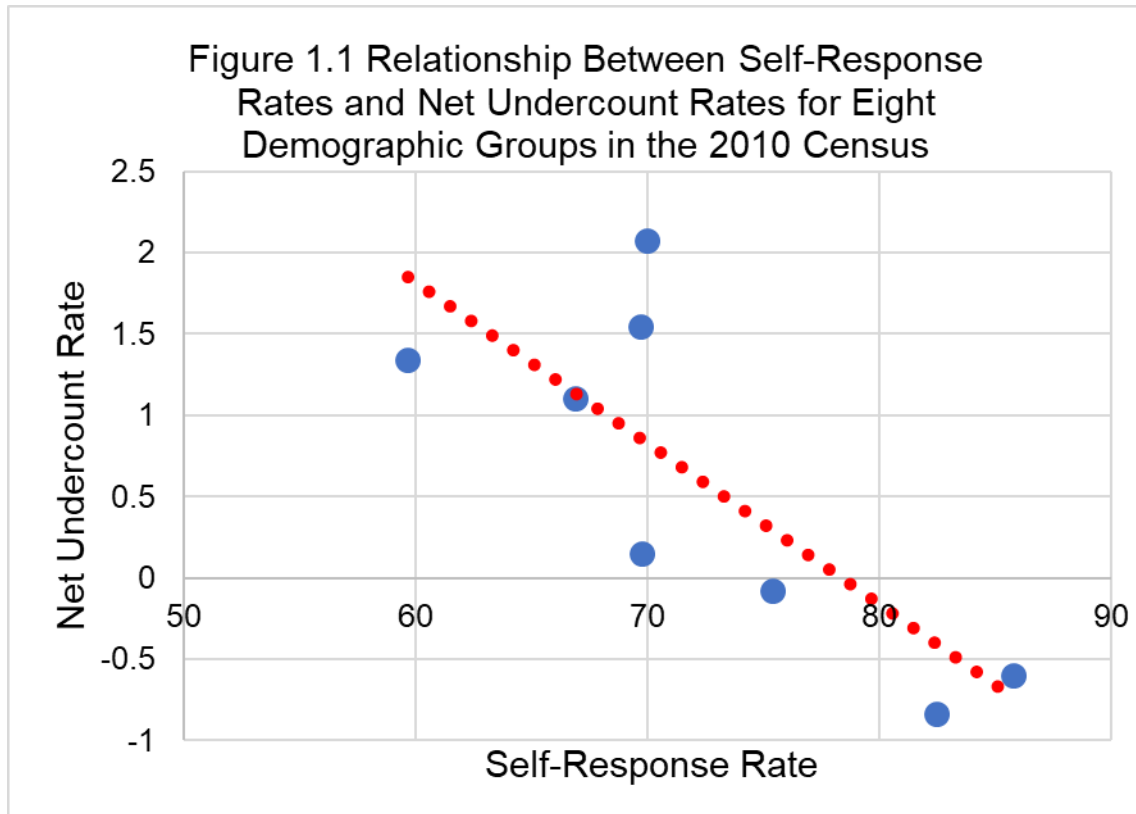
A. Examination of Data from the 2010 Census

28. Table 1.1 shows the self-response rates and net undercount rates for eight demographic groups defined by race, Hispanic Origin, and tenure (i.e. owner or renter). These are the only demographic groups where I could find all three measures (self-response, net undercount, and omissions rates) in consistently classified groups. The data for the self-response rates and undercount rates for these eight demographic groups were compiled by the Census Bureau after the 2010 Census. *See 2010 Census Mail Response/Return Rates Assessment Report*, 2012 (PX-341), at Table 10, and PX-267 at Table 7. The correlation coefficient between the self-response rate and the net undercount rate for the eight groups shown in Table 1.1 is -0.79. This correlation is statistically significantly different than zero at a 90 percent confidence level. This correlation is also statistically significantly different than zero at a 95 percent confidence level. This correlation means groups with lower self-response rates have higher net undercount rates.

29. The data in Table 1.1 are very consistent. All the demographic groups that have higher than average self-response rates have net overcounts and all groups with lower than average self-response rates have net undercounts. Asians essentially have self-response rates the same as the total population and essentially have a net undercount rate of zero.

Table 1.1 Self-Response Rates, and Net Undercount Rates for Demographic Groups in the 2010 Census		
	Self-Response Rates (Mail Return Rates)	Net Undercount Rates
Total	75.8	-0.01
White Alone	82.5	-0.84
Black Alone	70.0	2.07
American Indian and Alaskan Native Alone	69.8	0.15
Asian Alone	75.4	-0.08
Native Hawaiian or Pacific Islander Alone	59.7	1.34
Hispanic	69.7	1.54
Population in Owner-Occupied Housing Units	85.8	-0.60
Population in Renter- Occupied Housing Units	66.9	1.1

30. The relationship is shown graphically in Figure 1.1. The red dotted line shown in Figure 1.1 (and all other figures) is the trend line that reflects the statistical relationship between the two measures shown in the Figure. The closer the points are to the line, the higher the



correlation.

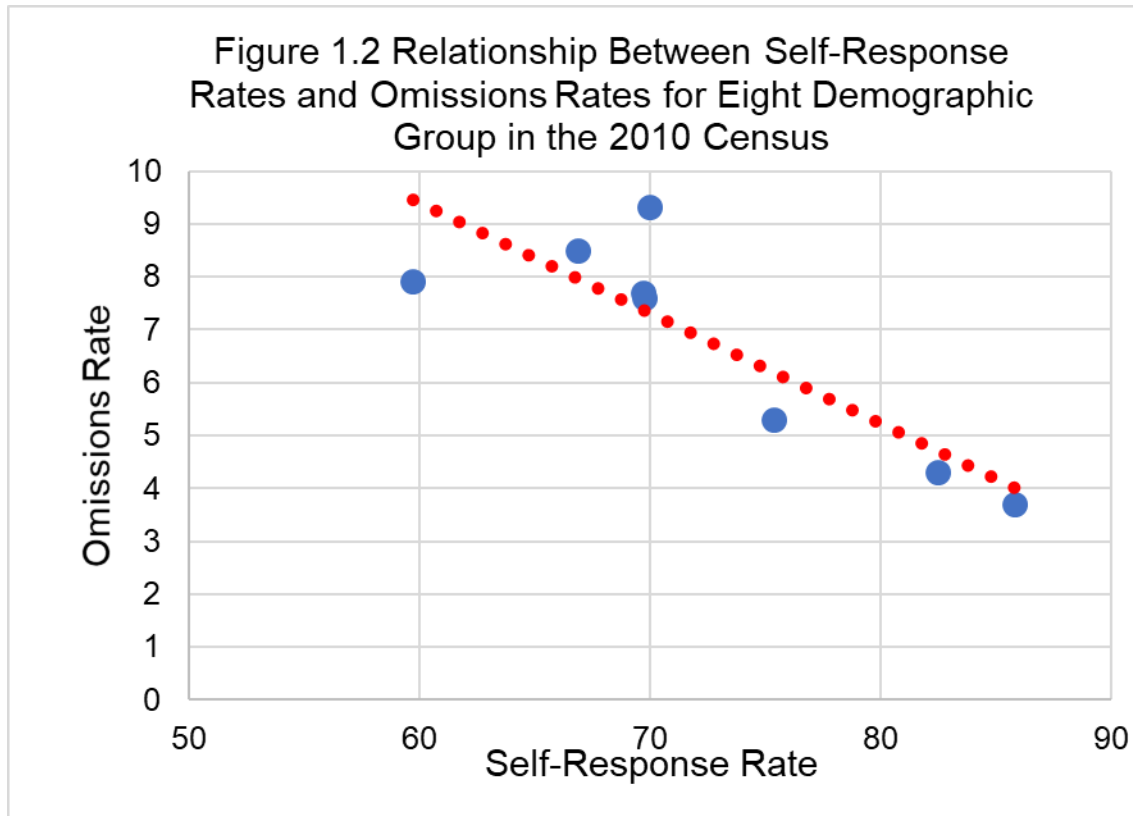
31. Table 1.2 shows 2010 Census self-response rates and omissions rates for the same eight demographic groups shown in Table 1.1. The omissions rates for these eight demographic groups were also compiled by the Census Bureau. *See U.S. Census Bureau, 2010 Components of Census Coverage for Race Groups and Hispanic Origin by Age, Sex, and Tenure in the United States, 2012.* The correlation coefficient between the self-response rates and the omissions rates for the eight groups in Table is -0.86 and it is statistically significantly different from zero at a 90 percent confidence level. The correlation means groups with lower self-response rates have

higher omissions rates.

Table 1.2 Self-Response Rates and Omission Rates for Demographic Groups in the 2010 Census		
	Self-Response Rates (Mail Return Rates)	Omission rates
Total	75.8	5.3
White Alone	82.5	4.3
Black Alone	70.0	9.3
American Indian and Alaskan Native Alone	69.8	7.6
Asian Alone	75.4	5.3
Native Hawaiian or Pacific Islander Alone	59.7	7.9
Hispanic	69.7	7.7
Population in Owner-Occupied Housing Units	85.8	3.7
Population in Renter- Occupied Housing Units	66.9	8.5

32. All the groups that have higher than average self-response rates have below average omissions rates and all the groups with lower than average self-response rates have higher than average omissions rates as shown in Table 1.2. Again, Asians essentially have self-response rates that are the same as that total population and have an omissions rate exactly equal to the total population.

33. The correlation can be seen graphically in Figure 1.2. This figure shows that groups with lower self-response rates have higher omissions rates.



34. Table 1.3 shows the self-response rates and net undercount rates for states along with the District of Columbia. These state self-response rates and net undercount rates were compiled by the Census Bureau after 2010 Census. *See State Mail Return Rates, 2010 (PX-348), and PX-267 at Table 14.* Note that none of the net undercount rates in Table 1.3 are statistically significantly different from zero. *See 2010 Census Coverage Measurement Estimation Report: Components of Census Coverage for the Household Population in the United States, 2012 (PX-338), at Table 5.* It is as if all the net undercount measures had the same value for all the states. This is important because when there is little variation in one of the measures in a relationship, it is more difficult to determine the extent of any correlation. Moreover, the sample size for each

state in the 2010 Post-Enumeration Survey was smaller than the Post-Enumeration Survey in 2000. The smaller sample size lead to less precise state estimates which contributed to the lack of statistical significance for this correlation. The lack of precision based on a small sample size along with little variation among states, means it is difficult to detect real differences between states.

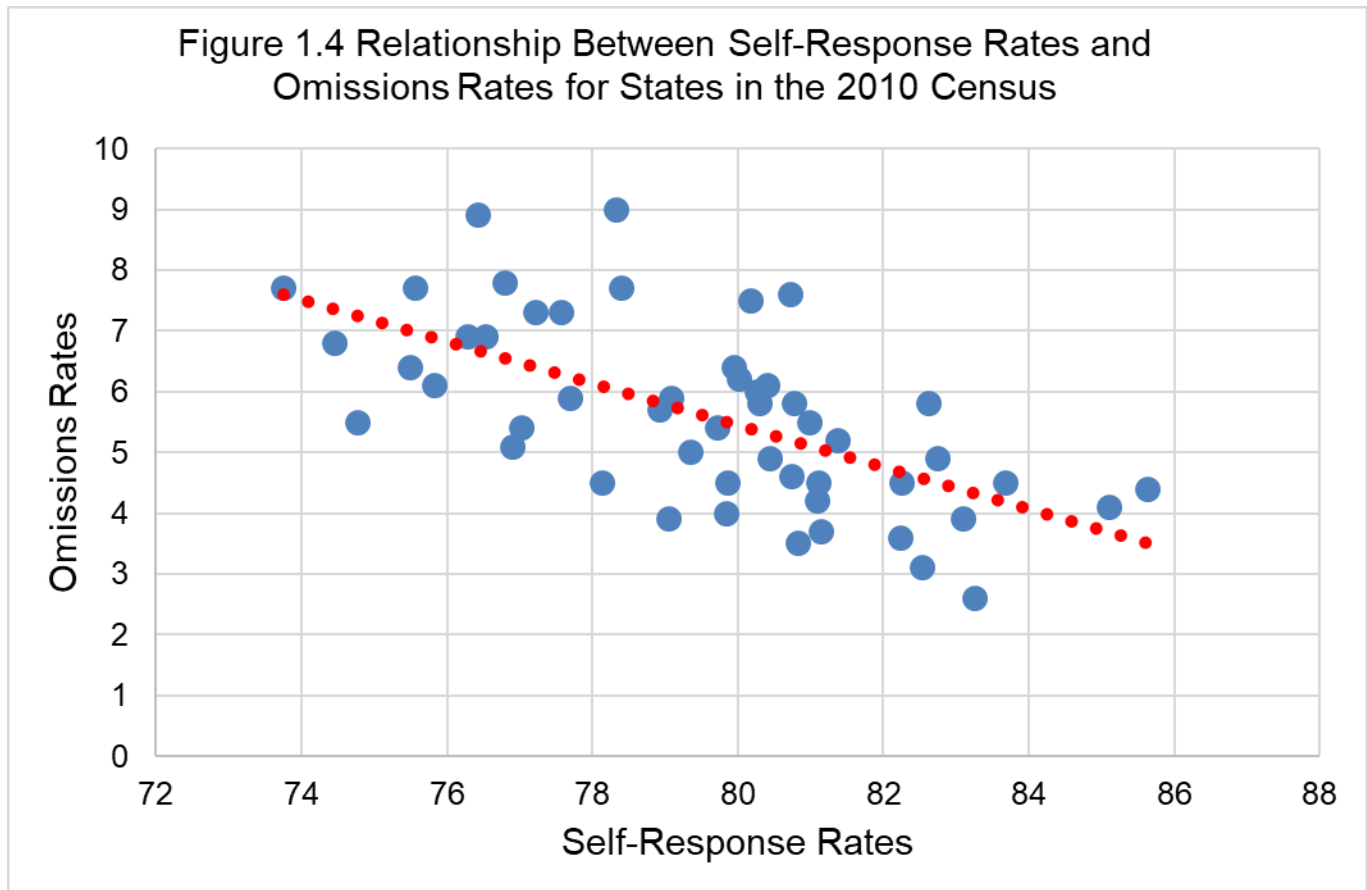
35. Given the lack of measurable variation in the state-level net undercount rates in the 2010 Census, correlation with self-response rates is expected to be low. Indeed, the correlation between self-response rates and net undercount rates across the states is 0.06, which is not statistically significantly different from zero. Given the low correlation between self-response rates and net undercount rates for states, the data are not shown graphically.

Table 1.3. 2010 Census Self-Response Rates and Net Undercount Rates by State						
State	Self-Response Rates (Mail Return Rate)	Net Undercount Rate		State	Self-Response Rates (Mail Return Rate)	Net Undercount Rate
Alabama	78.4	0.13		Montana	80.4	-0.65
Alaska	74.8	-0.85		Nebraska	82.5	-0.54
Arizona	77.6	-0.42		Nevada	76.3	-0.04
Arkansas	77.0	-0.41		New Hampshire	79.4	0.6
California	76.9	0.26		New Jersey	78.1	-0.36
Colorado	79.1	-0.29		New Mexico	73.8	-0.16
Connecticut	79.1	-0.45		New York	75.8	-0.79
Delaware	80.0	0.55		North Carolina	80.7	0.52
District of Columbia	78.3	2.23		North Dakota	83.1	0.09
Florida	80.2	0.45		Ohio	80.8	-0.83
Georgia	77.2	0.91		Oklahoma	75.5	-1.08
Hawaii	76.8	-0.44		Oregon	79.8	0.02
Idaho	82.6	-0.03		Pennsylvania	82.3	0.14
Illinois	80.7	-0.48		Rhode Island	77.7	-0.81
Indiana	82.2	-0.67		South Carolina	81.4	0.41
Iowa	83.3	-0.28		South Dakota	82.7	0.1
Kansas	81.2	-0.67		Tennessee	80.3	0.12
Kentucky	81.0	-0.13		Texas	76.5	0.97
Louisiana	74.5	-0.38		Utah	80.4	-0.48
Maine	81.1	0.65		Vermont	79.7	1.29
Maryland	80.3	0.94		Virginia	80.8	0.57
Massachusetts	78.9	-0.52		Washington	79.9	-0.1
Michigan	83.7	-0.66		West Virginia	75.6	-1.43
Minnesota	85.6	-0.56		Wisconsin	85.1	-0.17
Mississippi	76.4	0.24		Wyoming	79.9	-0.51
Missouri	81.1	-0.66		Total	75.8	-0.01

36. Table 1.4 shows self-response rates and omissions rates for states and the District of Columbia in the 2010 Census. These self-response and omissions rates were compiled by the Census Bureau. *See* PX-348 and PX-267 at Table 14. While the sample size for each state in the 2010 Post- Enumeration Survey was smaller than the 2000 Post-Enumeration Survey, the omissions rates varied a lot more across states and the correlation between self-response rates and omissions rates was statistically significant. That correlation was calculated at -0.63, and it is statistically significantly different from zero at a 90 percent confidence level. The correlation is also statistically significantly different than zero at a 95 percent confidence level. States that have lower self-response rates have higher omissions rates.

Table 1.4. 2010 Census Self-response Rates and Omissions Rates by State						
State	Self-Response Rates (Mail Return Rate)	Omissions Rate		State	Self-Response Rates (Mail Return Rate)	Omissions Rate
Alabama	78.4	7.7		Montana	80.4	6.1
Alaska	74.8	5.5		Nebraska	82.5	3.1
Arizona	77.6	7.3		Nevada	76.3	6.9
Arkansas	77.0	5.4		New Hampshire	79.4	5.0
California	76.9	5.1		New Jersey	78.1	4.5
Colorado	79.1	5.9		New Mexico	73.8	7.7
Connecticut	79.1	3.9		New York	75.8	6.1
Delaware	80.0	6.2		North Carolina	80.7	7.6
District of Columbia	78.3	9.0		North Dakota	83.1	3.9
Florida	80.2	7.5		Ohio	80.8	3.5
Georgia	77.2	7.3		Oklahoma	75.5	6.4
Hawaii	76.8	7.8		Oregon	79.8	4.0
Idaho	82.6	5.8		Pennsylvania	82.3	4.5
Illinois	80.7	4.6		Rhode Island	77.7	5.9
Indiana	82.2	3.6		South Carolina	81.4	5.2
Iowa	83.3	2.6		South Dakota	82.7	4.9
Kansas	81.2	3.7		Tennessee	80.3	5.8
Kentucky	81.0	5.5		Texas	76.5	6.9
Louisiana	74.5	6.8		Utah	80.4	4.9
Maine	81.1	4.2		Vermont	79.7	5.4
Maryland	80.3	6.0		Virginia	80.8	5.8
Massachusetts	78.9	5.7		Washington	79.9	4.5
Michigan	83.7	4.5		West Virginia	75.6	7.7
Minnesota	85.6	4.4		Wisconsin	85.1	4.1
Mississippi	76.4	8.9		Wyoming	79.9	6.4
Missouri	81.1	4.5		Total	75.8	5.3

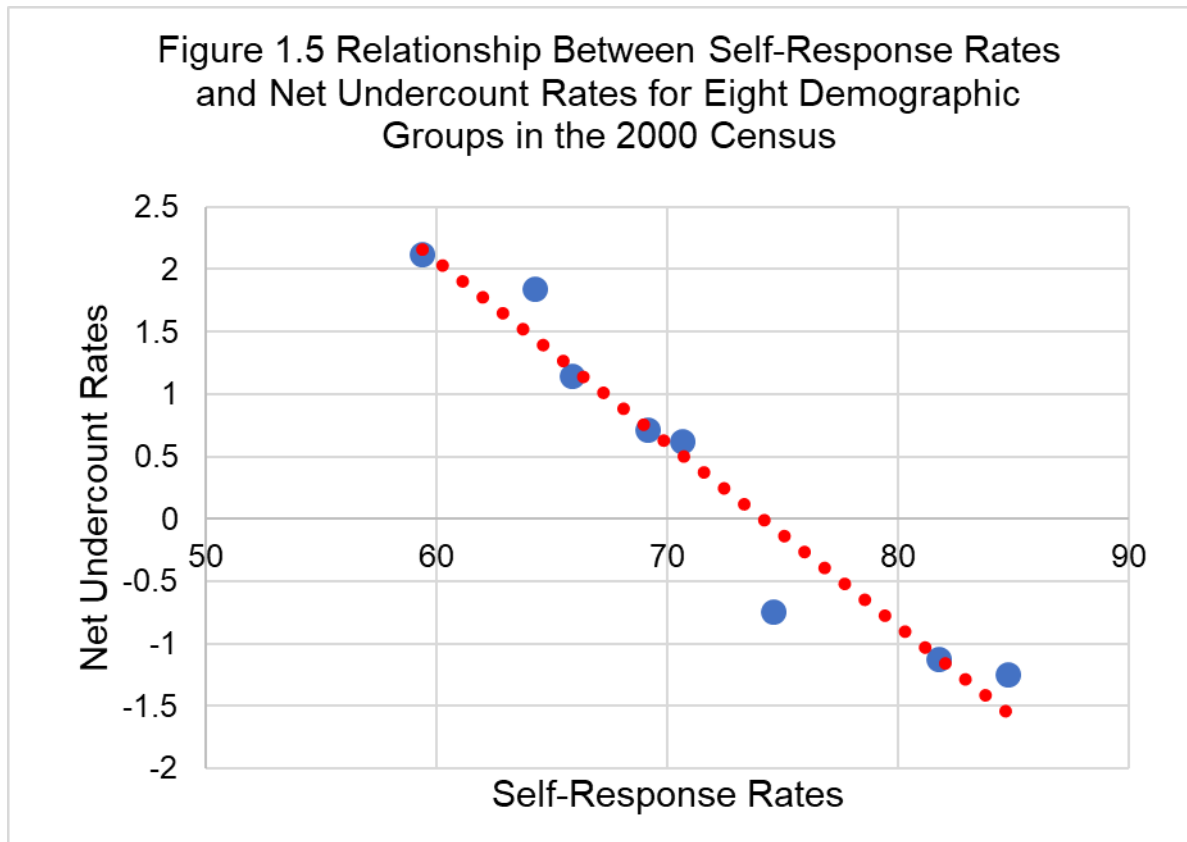
37. Figure 1.4 shows the relationship between self-response rates and omissions rates for states in the 2010 Census graphically. Figure 1.4 shows that states with lower self-response rates have higher omissions rates.



38. Table 1.5 shows self-response rates and net undercount rates from the 2000 Census for eight demographic groups. The data were calculated by the Census Bureau shortly after the 2000 Census. See *Census 2000 Mail Return Rates*, 2003 (PX-342), Tables 10, 12 and 16, and *DSSD A.C.E. Revision II Memorandum Series #PP-54* (PX-349), at Table 1. Note that the racial groups are not defined exactly the same in the two Census Bureau reports from which the data were taken but they are very similar. This is a minor point and unlikely to impact the correlation. The correlation coefficient between self-response rates and net undercount rates in Table 1.5 is -0.97, which is very high and statistically significantly different from zero at a 90 percent confidence level. This correlation coefficient is also statistically significantly different than zero at a 95 percent confidence level. This means that demographic groups that have low self-response rates have high net undercounts.

Table 1.5 Self-Response Rates and Net Undercount Rates in the 2000 Census for Eight Demographic Groups			
Groups	Self-Response Rates (Mail Return Rate)	Groups	Net Undercount Rates (A.C.E. Revision II)
White Alone	81.8	Non-Hispanic White	-1.13
Black Alone	64.3	Non-Hispanic Black	1.84
Asian Alone	74.6	Non-Hispanic Asian	-0.75
Pacific Islander Alone	59.4	Hawaiian or Pacific Islander	2.12
Hispanic	69.2	Hispanic	0.71
American Indian Alone	70.7	AIAN Off Reservations	0.62
Owner-Occupied	84.8	Homeowner	-1.25
Renter-Occupied	65.9	Renter	1.14

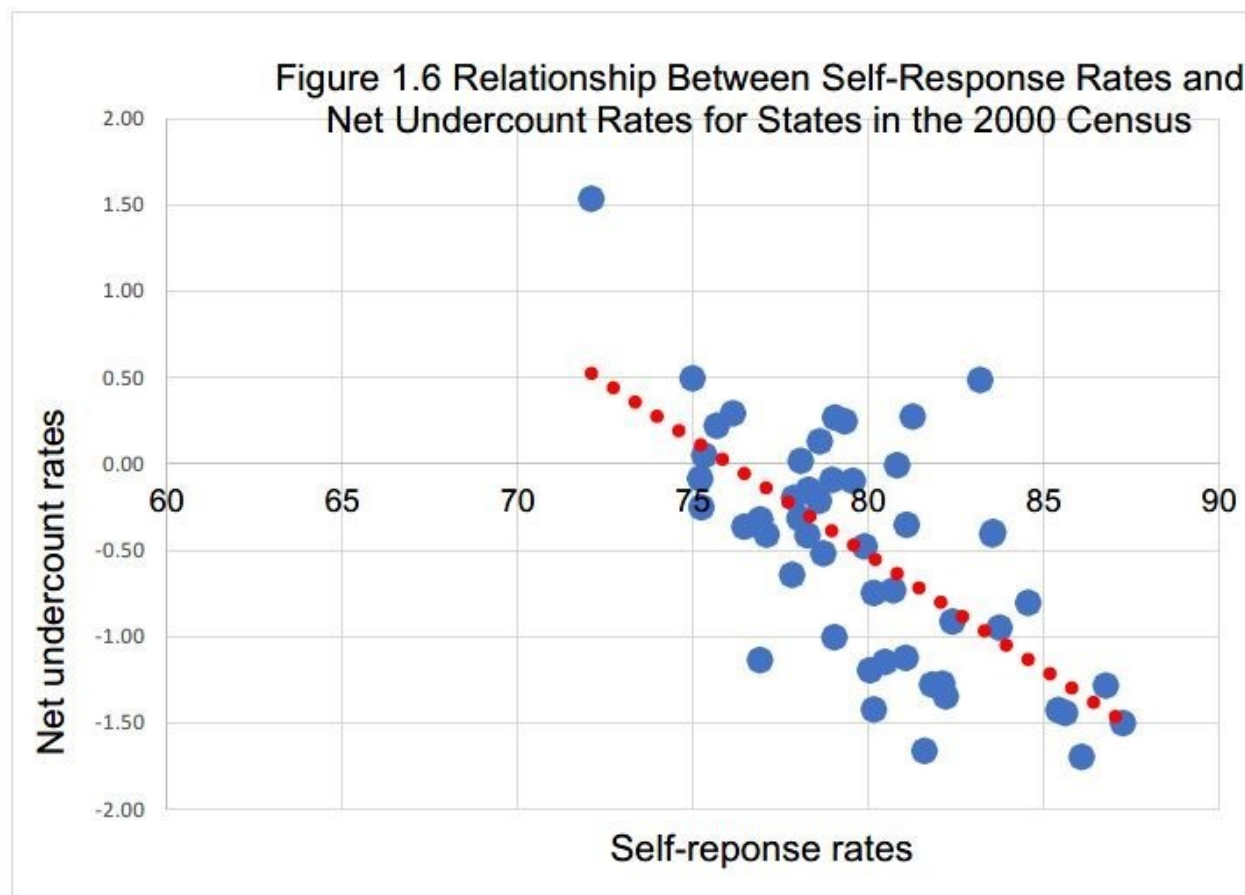
39. The relationship between self-response rates and net undercount rates is shown graphically in Figure 1.5. This figure shows that groups that have lower self-response rates have higher net undercount rates.



40. Table 1.6 shows the self-response rates and net undercount rates for states along with the District of Columbia in the 2000 Census. These self-response and net undercount rates were compiled by the Census Bureau. See PX-348 and *A.C.E. Revision II – Adjusted Data for States, Counties, and Places*, 2003 (PX-350), at Table 1. The correlation between self-response rates and net undercount rates across the states is -0.66, which is statistically significantly different from zero at a 90 percent confidence level. This correlation is also statistically significantly different than zero at a 95 percent confidence level. Groups with low self-response rates have high net undercount rates. To the best of my knowledge state omissions rates for the 2000 Census were not readily available to the public.

Table 1.6 Census 2000 Self-Response Rates and Undercount Rates for States						
State	Self-Response Rates (Mail Return Rates)	Net Undercount Rates		State	Response Rates (Mail Return)	Net Undercount Rates
Alabama	76.9	-0.34		Montana	83.2	0.49
Alaska	76.1	0.29		Nebraska	84.6	-0.81
Arizona	76.9	-0.32		Nevada	75.0	0.50
Arkansas	79.0	-0.09		New Hampshire	80.5	-1.15
California	78.6	0.13		New Jersey	78.7	-0.52
Colorado	80.8	-0.01		New Mexico	78.1	0.02
Connecticut	80.2	-0.75		New York	75.2	-0.25
Delaware	78.0	-0.32		North Carolina	78.3	-0.15
District of Columbia	72.1	1.54		North Dakota	85.4	-1.43
Florida	77.8	-0.64		Ohio	82.1	-1.27
Georgia	79.1	0.27		Oklahoma	77.9	-0.20
Hawaii	75.7	0.22		Oregon	81.1	-0.35
Idaho	83.5	-0.41		Pennsylvania	82.4	-0.91
Illinois	80.2	-1.42		Rhode Island	76.9	-1.14
Indiana	81.6	-1.66		South Carolina	76.5	-0.36
Iowa	85.6	-1.44		South Dakota	86.8	-1.28
Kansas	81.8	-1.28		Tennessee	77.1	-0.41
Kentucky	79.9	-0.48		Texas	75.3	0.05
Louisiana	75.2	-0.09		Utah	79.6	-0.10
Maine	80.1	-1.20		Vermont	81.1	-1.12
Maryland	79.3	0.25		Virginia	81.3	0.27
Massachusetts	79.0	-1.00		Washington	78.6	-0.21
Michigan	83.7	-0.95		West Virginia	80.7	-0.73
Minnesota	86.1	-1.70		Wisconsin	87.3	-1.50
Mississippi	78.3	-0.41		Wyoming	83.6	-0.39
Missouri	82.2	-1.35		United States	78.4	0.48

41. This relationship is shown graphically in Figure 1.6. This figure shows that states that have lower self-response rates have high net undercount rates.



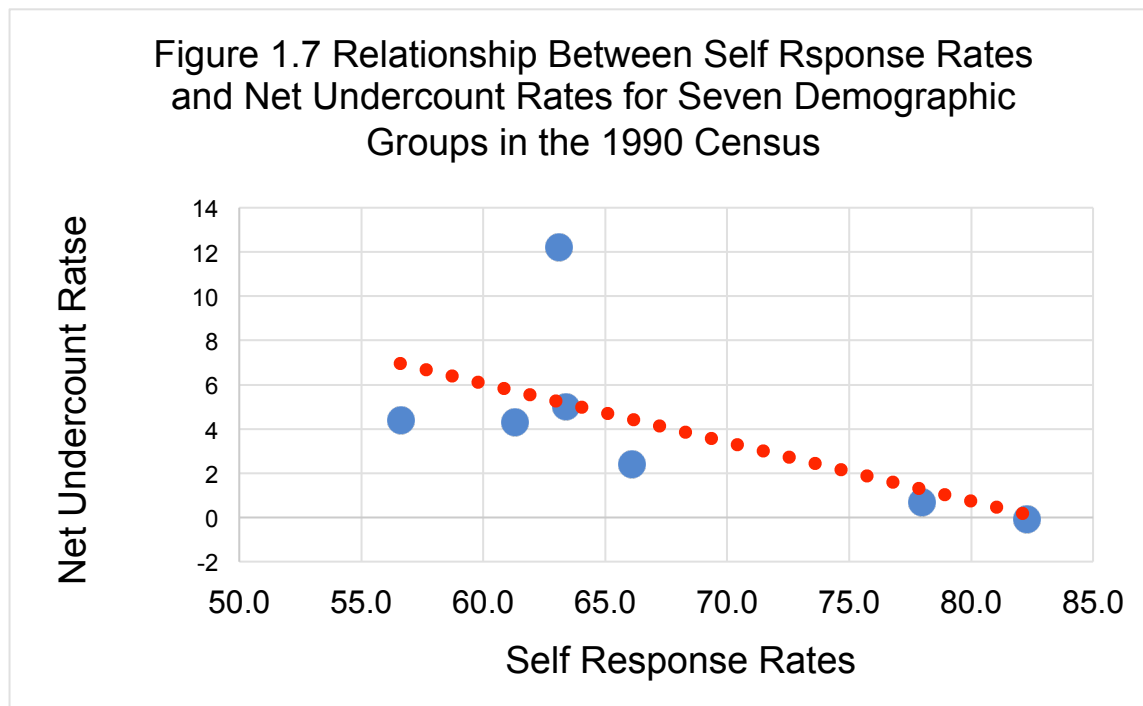
B. Examination of Data from the 1990 Census

42. The only self-response rates available for states in the 1990 Census were Mail Response Rates. Mail Response Rates are slightly different than the Mail Return Rates (as explained in Paragraph 12), but both are measures of self-response used by the Census Bureau.

43. Table 1.7 shows self-response rates and net undercount rates in the 1990 Census for seven demographic groups. The Census Bureau compiled these self-response rates and net undercount rates shortly after the 1990 Census. *See* Word, *Who Responds? Who Doesn't?: Analyzing Variation in Mail response Rates During the 1990 Census*, 1997 (PX-343), at Table 2.0, and Hogan and Robinson, *What the Census Bureau's Coverage Evaluation Programs Tell Us About Different Undercounts*, 1993 (PX-351), at Table 3. The correlation between the self-response rates and the net undercount rates is -0.61, which is statistically significantly different from zero at a 90 percent confidence level. The correlation indicates that groups with lower self-response rates have higher net undercount rates.

Table 1.7 1990 Census Self-Response Rates and Net Undercount Rates for Seven Demographic Groups			
	Self-Response Rates (Mail Responses Rates)		Net Undercount Rate
Non-Hispanic White	78.0	Non-Hispanic White	0.7
Black	56.6	Black	4.4
American Indians, Eskimo and Aleut	63.1	American Indian, Eskimo and Aleut	12.2
Asian and Pacific Islanders	66.1	Asians and Pacific Islander	2.4
Hispanic Origin	63.4	Hispanic Origin	5.0
Owners	82.3	Owners	-0.1
Renters	61.3	Renters	4.3

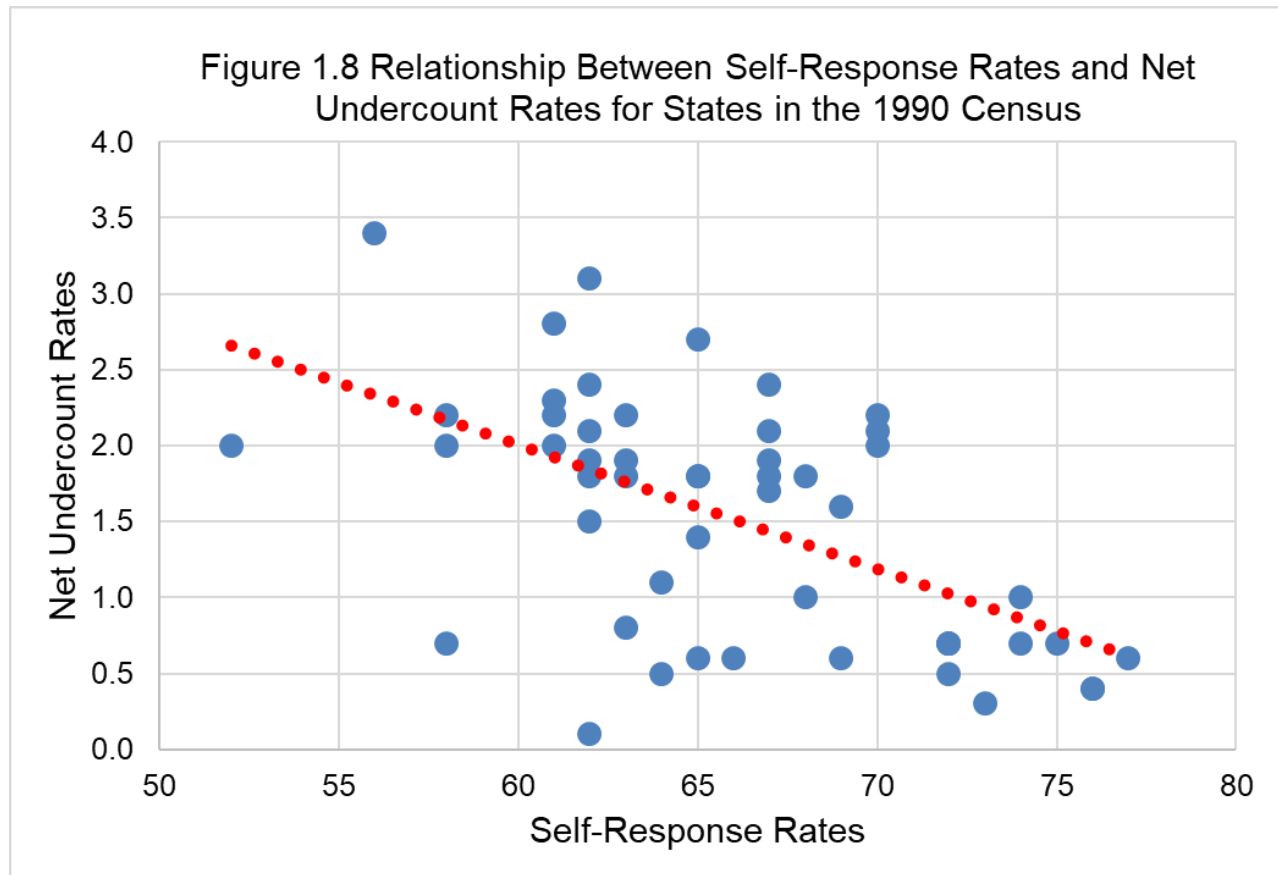
44. This relationship is shown graphically in Figure 1.7. This figure shows that groups that have low self-response rates have higher net undercount rates.



45. Table 1.8 shows 1990 Census self-response rates and net undercount rates for states and the District of Columbia. These self-response and net undercount rates were compiled by the Census Bureau. *See 1990 Mail response Rates by 1990 Geography Boundaries* (PX-352), and PX-267 at Table 7. The correlation coefficient between self-response rates and net undercount rates, shown in Table 1.8 is -0.56, which is statistically significantly different from zero at a 90 percent confidence level. This correlation coefficient is also statistically significantly different than zero at a 95 percent confidence level. This means that states with lower self-response rates have higher net undercount rates. To the best of my knowledge, state omissions rates from the 1990 Census were not readily available to the public.

Table 1.8 1990 Census Self-Response Rates and Net Undercount Rates for States						
State	Self-Response Rates (Mail Response Rates)	1990 Net Undercount of Persons		State	Self-Response Rates (Mail Response Rates)	1990 Net Undercount of Persons
Alabama	62.0	1.8		Montana	67.0	2.4
Alaska	52.0	2.0		Nebraska	74.0	0.7
Arizona	62.0	2.4		Nevada	61.0	2.3
Arkansas	65.0	1.8		New Hampshire	63.0	0.8
California	65.0	2.7		New Jersey	65.0	0.6
Colorado	67.0	2.1		New Mexico	62.0	3.1
Connecticut	66.0	0.6		New York	62.0	1.5
Delaware	68.0	1.8		North Carolina	63.0	1.9
District of Columbia	56.0	3.4		North Dakota	72.0	0.7
Florida	61.0	2.0		Ohio	75.0	0.7
Georgia	63.0	2.2		Oklahoma	63.0	1.8
Hawaii	62.0	1.9		Oregon	67.0	1.9
Idaho	70.0	2.2		Pennsylvania	73.0	0.3
Illinois	68.0	1.0		Rhode Island	62.0	0.1
Indiana	72.0	0.5		South Carolina	58.0	2.0
Iowa	76.0	0.4		South Dakota	74.0	1.0
Kansas	72.0	0.7		Tennessee	65.0	1.8
Kentucky	69.0	1.6		Texas	61.0	2.8
Louisiana	58.0	2.2		Utah	67.0	1.7
Maine	58.0	0.7		Vermont	64.0	1.1
Maryland	70.0	2.1		Virginia	70.0	2.0
Massachusetts	64.0	0.5		Washington	67.0	1.8
Michigan	72.0	0.7		West Virginia	65.0	1.4
Minnesota	76.0	0.4		Wisconsin	77.0	0.6
Mississippi	62.0	2.1		Wyoming	61.0	2.2
Missouri	69.0	0.6		U.S. Total		1.6

46. Figure 1.8 shows the relationship between state 1990 Mail Response Rates and net undercount rates graphically. This figure shows that states that have low self-response rates have high net undercount rates.



C. Summary of Relationship between Self-Response Rates and Census Accuracy

47. Table 1.9 summarizes the correlations between self-response rates and census accuracy for the past three Decennial Censuses. Based on empirical statistical relationships seen in the past three Decennial Censuses, the Census Bureau's expected decrease of at least 5.1 percentage points in the self-response rates for households with at least one noncitizen due to the addition of the citizenship question will increase the net undercount and omissions rates for people living in those households in the 2020 Census.

48. Of the eight correlations shown in Table 1.9, all but one of the correlations were

in the predicted direction and statistically significant. There was a reasonable explanation for the one correlation that was low, in the opposite direction and not statistically significant. *See* Paragraphs 34-35. Based on my 40 plus years of experience as a professional data analyst I would call the correlation coefficients in Table 1.9 (except for 0.06) moderate to high by social science standards.

Table 1.9 Summary of Statistical Relationships between Census Self-Response Rates and Census Accuracy (Net Undercount Rates and Omissions Rates)	
Correlation between Self-Response Rates and:	Correlation Coefficient
2010 Eight Demographic Groups Net Undercount	-0.78
2010 Eight Demographic Groups Omissions	-0.86
2010 States Net Undercount Rates	0.06
2010 States Omissions Rates	-0.63
2000 Eight Demographic Groups	-0.97
2000 States Net Undercount Rates	-0.66
1990 Seven Demographic Groups	-0.61
1990 States Net Undercount Rates	-0.56
Note correlations in BOLD are statistically significant at the .10 level or higher.	

49. Social scientists typically look for four elements to show causation. First, that the causal agent (referred to as the independent variable by scientists) occurs prior in time to the thing that it is causing (referred to as the dependent variable by scientists); second, that there is an association or correlation between the causal agent and the thing being caused; third, that intervening mechanisms linking the independent variable and the dependent variable can be

clearly specified; and finally that other potential explanations have been controlled.

50. My analysis satisfies three out of four of these elements. Self-response occurs prior in time to net undercounting, the self-response rate is moderately to highly correlated with net undercounting, and the intervening mechanism is the fact that groups with lower self-response rates have a higher share of their population counted in the NRFU operation which generates less accurate data. The only element my analysis does not address is need to control for other potential explanations. The inability to control for all other potential explanations is common in social science research because there are legal and ethical restrictions on how much people can be manipulated for research purposes. The way to control for all other possible explanations is through a randomized control trial (“RCT”). However, the Census Bureau, which is best positioned to conduct an RCT, has not conducted any such RCT measuring the relationship between self-response and undercounting. My analysis evaluates the relationship between self-response rates and census accuracy with best available data, and shows there is a strong robust relationship between self-response and Census accuracy; namely groups that have lower self-response rates have higher net undercount and omissions rates.

51. I did not attempt to use this analysis to predict the exact increase in net undercounts and omissions for the 2020 Census. The exact predictions are not the point. The preponderance of evidence clearly shows there would be an increase in net undercounts and omissions because of the addition of the question on citizenship. The point is the preponderance of evidence from past empirical relationships indicates the decrease in self-response rates for households with one or more noncitizens will lead to an increase in net undercounts and omissions. The important point is that there will be a differential impact on Census accuracy, not the magnitude of that differential. The magnitude of the correlations varies from one Census to

the next, but they are consistent in showing a negative correlation between self-response rates and census accuracy.

52. The consistency of the correlation (7 out of 8 observations) across multiple Censuses, demographic groups and states, is illustrative of a consistent relationship over time. The analysis demonstrates a clear pattern in the relationship between self-response and undercount rates. While there is some uncertainty in these data (as with all data), uncertainty typically reduces the likelihood of finding a correlation. But here, my analysis demonstrates a correlation in spite of this uncertainty, providing further proof that the relationship is real.

53. The empirical relationship between self-response rates and Census accuracy (net undercounts and omissions) has been recognized by others. The Census Bureau Task Force on the Undercount of Young Children concluded, “Research suggests that areas with lowest levels of cooperation have higher levels of coverage and nonresponse error.” *See Final Task Force Report* (PX-346), at ii. In a Census Bureau Working Paper Word notes: “...response rates and net undercount rates may be causally linked...” *See* PX-343 at 1.

54. The connection between self-response rates and Census accuracy is underscored by the Census Bureau’s decision to use a self-response related measure to identify Hard-to-Count areas in the 2020 Census. The Census Bureau’s Low-Response Score, developed by Erdman and Bates, is based on the Mail Return Rates in the 2010 Census. In describing the Low Response Score the Census Bureau has stated that “[t]his score identifies Block Groups and Tracts whose characteristics predict low Census Mail Return Rate and are highly correlated (negatively) with Census and survey participation.” *See* PX-345 at 4. The implicit association here is that areas where self-response rates are low are more difficult to enumerate.

55. It is easy to understand why the relationship between low self-response rates and

Census accuracy exists. Households that do not self-respond end up in the NRFU pool where the Census Bureau may have to rely on a proxy response. A recent Census Bureau White Paper addressing the potential impact of adding a citizenship question to the Decennial Census indicates they expect 21 percent of households with at least one noncitizens who do not self-respond will end up being counted by proxy responses. *See* Brown et al., *Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census*, 2018, at 42. The Census Bureau White Paper also shows that the quality of data from the self-response portion of the Census is much more accurate than the data collected in NRFU and/or proxy response portion. *See Id.* at Table 42. The data showed that 97.3 percent of the responses from the Mailout/Mail back portion of the Census were correct, as compared to just 70.2 percent of those from the NRFU proxy responses.

IV. CONCLUSION

56. The Census Bureau stipulates that the addition of the citizenship question on the 2020 Census questionnaire will lower self-response rates for households with one or more noncitizens. The preponderance of empirical evidence from the 2010, 2000 and 1990 Census shows lower self-response rates lead to higher net undercount rates and omissions rates.

57. The Census Bureau has asserted that the people who do not self-respond will be picked up in the NRFU operation. However, there are two flaws in this assertion. Among other things, NRFU does not address people left off census questionnaires for households that self-respond. It assumes that everyone in a responding household is included on the census questionnaire and it ignores the likely differential impact of noncitizens in this stage of the data collection process. This is inconsistent with evidence from the 2010 Census. For example, a study by U.S. Census Bureau based on matching young children found in the Post-Enumeration

Survey to records from the 2010 Census shows that more than 84 percent of young children missed in the Census lived in households that were included in the census (or at least their household was included in the Master Address File and they received a questionnaire). See *Assessing Net Coverage Error for Young Children in the 2010 U.S. Decennial Census*, PX-339.

58. Dr. Abowd suggests in his expert report that all the people in nonresponding households will be captured in the NRFU portion of the Census. But the Non-Response Follow Up operation has not mitigated low response rates in the past so there is no reason to believe it will do so in the 2020 Census. Non-Response Follow Up has not worked to perfection in the past, and there is no reason to believe that it will solve the issue of low self-response rates among households with one or more noncitizens. Data collected in the NFRU portion of the Census is less accurate than that collected in the self-response portion. Decreased self-response translates into more omissions and increased undercount of those populations with reduced self-response rates.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on: 11/6, 2018


William O'Hare